

[12] 发明专利申请公开说明书

[21] 申请号 00130367.8

[43] 公开日 2002 年 6 月 12 日

[11] 公开号 CN 1353367A

[22] 申请日 2000.11.2 [21] 申请号 00130367.8

[71] 申请人 北京算通数字技术研究中心有限公司

地址 100080 北京市中关村科学院南路 8 号

[72] 发明人 高文 庞斌

[74] 专利代理机构 北京三友知识产权代理有限公司

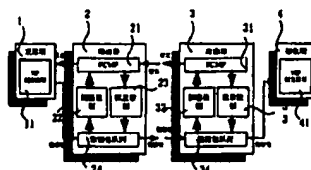
代理人 刘芳

权利要求书 4 页 说明书 7 页 附图页数 4 页

[54] 发明名称 一种基于随机早期检测的逐节拥塞控制方法

[57] 摘要

一种基于随机早期检测的逐节拥塞控制方法,其包括流量控制信息协议、基于该协议的拥塞检测方法、路由器流量控制方法和主机流量控制方法;网络各节点利用改进的随机早期检测算法检测拥塞状态,在拥塞发生之前向邻近的上游节点发送流量控制信息;当节点收到流量控制信息后,实施流量控制,避免和减轻了网络数据的拥塞,克服了现有网络拥塞控制的缺陷;同时,由于各网络节点不用保存每条连接的状态信息,使其具有良好的可扩展性。



ISSN 1008-4274

知识产权出版社出版

BEST AVAILABLE COPY

1、一种基于随机早期检测的逐节拥塞控制方法，其特征在于：它包括流量控制信息协议、基于该协议的拥塞检测方法、利用该协议的路由器流量控制方法和主机流量控制方法。

2、如权利要求 1 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：该流量控制信息协议为一位于 TCP/IP 协议族的 IP 层和数据链路层之间、用于相邻路由器之间及路由器与主机之间交换拥塞状态信息的流量控制信息协议报文格式。

3、如权利要求 2 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：该流量控制信息协议报文格式设有：用于验证发送方和接收方的流量控制信息协议报文格式一致性的版本号域、流量控制信息协议报文长度域、流量控制信息协议报文类型域、用于表示路由器类型的代码域、源 IP 地址域、目的 IP 地址域、源端口号域、目的端口号域、用于保证流量控制信息协议报文数据的完整性的校验域及填充域。

4、如权利要求 3 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：流量控制信息协议报文类型包括恢复流量报文类型及降低流量报文类型。

5、如权利要求 3 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：路由器类型包括与发送端主机或接收端主机相连接的边缘路由器及仅与路由器相连接的内部路由器。

6、如权利要求 1 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：所述的基于流量控制信息协议的拥塞检测方法包括如下步骤：

a、将数据包队列的平均长度与预定义的队列平均长度的上限和下限作比较；

b、当数据包队列的平均长度在上、下限之间，则认为有可能发生拥塞；计算标记的概率，并根据该概率决定是否对将要入队的数据包做标记；否则转

步骤 d;

c、路由器向发送该已标记的数据包的上游节点发送“降低流量”流量控制信息协议报文;

d、当数据包队列的平均长度小于下限, 不对数据包作标记; 否则将该包

5 丢弃;

e、如果在此之前, 发送过“降低流量”的流量控制信息协议报文, 再发送“恢复流量”流量控制信息协议报文。

7、如权利要求 6 所述的基于随机早期检测的逐节拥塞控制方法, 其特征在于: 所述的拥塞检测是在数据包队列到达本节点时开始的。

10 8、如权利要求 6 所述的基于随机早期检测的逐节拥塞控制方法, 其特征在于: 所述的平均队列长度依照如下公式计算:

当队列不空时 ($q > 0$), 则: $avg \leftarrow (1 - W_q) \times avg + W_q \times q;$

否则: $m \leftarrow f(time - q_time);$

$avg \leftarrow (1 - W_q)^m \times avg;$

15 其中: avg = 平均队列长度;

$time$ = 当前时间;

q_time = 队列空闲时间的起始点;

W_q = 队列的权重;

q = 队列的实际长度;

20 $f(t)$ = 时间 t 的一个线性函数;

9、如权利要求 6 所述的基于随机早期检测的逐节拥塞控制方法, 其特征在于: 所述的标注概率依照如下公式和步骤计算:

将 $count$ 加 1;

$P_s \leftarrow P_{max} (avg - TH_{min}) / (TH_{max} - TH_{min})$

25 $P_s \leftarrow P_s / (1 - count \times P_s)$

其中: $count$ = 上一个“降低流量”流量控制信息协议报文发送后数据包

的分组数;

TH_{min} = 队列长度的下限;

TH_{max} = 队列长度的上限;

P_0 = 计算中间值;

5 P_{max} = P_0 的最大值;

P_i = 当前数据包分组进行标注的概率。

10、如权利要求 1 所述的基于随机早期检测的逐节拥塞控制方法, 其特征在于: 所述的路由器流量控制方法包括如下步骤:

a、路由器通过“漏桶算法”控制流量;

10 b、当收到“降低流量”流量控制信息协议报文时, 将漏桶容量减半;

c、每隔一设定的时间间隔将漏桶容量加 1 个长度的数据包, 但最大值不超过漏桶的最初容量;

d、当收到“恢复流量”流量控制信息协议报文时, 立即恢复漏桶的最初容量。

15 11、如权利要求 10 所述的基于随机早期检测的逐节拥塞控制方法, 其特征在于: 所述的漏桶算法为如下步骤:

a、以一定的速率产生令牌, 令牌桶满时, 终止令牌的产生;

b、当数据包到达时, 如果令牌桶中有令牌, 则取出一个令牌后发送数据包; 否则数据包放在缓冲区中等待令牌的出现;

20 c、当缓冲区满时, 丢弃到达的数据包。

12、如权利要求 1 所述的基于随机早期检测的逐节拥塞控制方法, 其特征在于: 所述的主机流量控制方法包括如下步骤:

a、主机根据由边缘路由器发来的流量控制信息协议报文中各域的信息查找对应 TCP 连接的套接字;

25 b、找到对应的套接字后, TCP 通过调用支持流量控制信息协议的拥塞控制过程来控制流量;

c、如果 TCP 在上一往返时间内没有降低流量，流量控制的请求将在本往返时间内被响应一次；

d、如果该流量控制请求是丢包引起的，TCP 调用原来的拥塞控制过程；

e、如果收到“降低流量”流量控制信息协议报文，TCP 将拥塞窗口和慢

5 启动临界值减半。

13、如权利要求 12 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：所述的往返时间为一个数据包从一台主机出发到达另一台主机然后返回所用的时间和。

10 14、如权利要求 12 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：所述的原来的拥塞控制过程为以 TCP 为核心的基于窗口技术的端到端控制方式。

15、如权利要求 12 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：所述的拥塞窗口为 TCP 在启动阶段使用的或在拥塞时期为减小流量而使用的窗口。

15 16、如权利要求 12 所述的基于随机早期检测的逐节拥塞控制方法，其特征在于：所述的慢启动临界值为：在慢启动阶段，拥塞窗口的增加方式由以指数增加方式改变为以线性增加方式时的拥塞窗口数量。

一种基于随机早期检测的逐节拥塞控制方法

5 本发明涉及一种基于随机早期检测(Random Early Detection)的逐节拥塞控制方法,属于计算机网络领域。

现在的互连网(Internet)已由以往的单一数据传送网发展成传送数据、语音、视频等多媒体信息的综合业务网。随着网络应用和用户的不断增加,Internet 经常发生拥塞,造成了数据传输延时、抖动等服务质量的下降,使
10 带宽、缓存等网络资源利用率降低,在极端的情况下,可导致整个网络崩溃。

目前在 Internet 上的拥塞控制方法是以 TCP 为核心的基于窗口技术的端到端控制。这种方法主要是以丢包作为拥塞发生的信号,当发送端确认已发出的数据包丢失后,通过减少拥塞窗口来降低发送速度,缓解拥塞。但是,由于 Internet 规模的不断增长,从网络节点发生拥塞(丢包)到发送端降低流量
15 需要较长的响应周期,而此时发送端仍在高速度发送数据,这会造成更大规模的拥塞,因此不能完全依靠所有的发送端控制拥塞。

近几年来出现的另一种方法是逐节拥塞控制。它通过利用每条连接的一些状态信息(如缓冲区的大小和发送速度等)在数据流经过的每个网络节点上实施拥塞控制。但这种方法要在网络节点中保存每条连接的一些状态信息,当连接数量增长时,存储容量和处理时间也随之增加,因而该方法在 Internet 主干网上没有很好的可扩展性。
20

本发明的主要目的在于提出一种基于随机早期检测的逐节拥塞控制方法,各网络节点用改进的随机早期检测算法检测拥塞状态,在拥塞发生之前向邻近的上游节点发送以流量控制信息协议(Flow Control Message Protocol, 简称
25 为 FCMP)规定格式的流量控制信息。

本发明的又一目的在于提出一种基于随机早期检测的逐节拥塞控制方法,当节点收到流量控制信息时,实施流量控制,避免和减轻拥塞,克服现有网络

拥塞控制的缺陷;各网络节点不用保存每条连接的状态信息,使其具有良好的可扩展性。

本发明的目的是通过以下的技术方案实现的:

一种基于随机早期检测的逐节拥塞控制方法,其包括流量控制信息协议、

- 5 基于流量控制信息协议的拥塞检测方法、路由器流量控制方法和主机流量控制方法。其中:

流量控制信息协议为一种报文格式,该报文在 TCP/IP 协议族的 IP 层和数据链路层之间传递,且用于在相邻路由器之间以及路由器与主机之间交换拥塞状态信息;该报文格式包括如下的信息域:用于验证发送方和接收方的流量控制信息协议报文格式一致性的版本号域、流量控制信息协议报文长度域、流量控制信息协议报文类型域、用于表示路由器类型的代码域、源 IP 地址域、目的 IP 地址域、源端口号域、目的端口号域、用于保证流量控制信息协议报文数据的完整性的校验域和填充域。上述的流量控制信息协议报文类型包括恢复流量报文类型和降低流量报文类型,路由器类型包括与发送端主机或接收端主机相连接的边缘路由器及仅与路由器相连接的内部路由器。

10

15

拥塞检测是在数据包队列到达本节点时开始进行的;其方法包括如下步骤:

- 1、将数据包队列的平均长度与预定义的队列平均长度的上限和下限作比较;
- 20 2、当数据包队列的平均长度在上、下限之间,则认为有可能发生拥塞;计算标注概率,并根据该概率决定是否对将要入队的数据包做标记;否则转步骤 4;
- 3、路由器向发送该已标记的数据包的上游节点发送“降低流量”流量控制信息协议报文;
- 25 4、当数据包队列的平均长度小于下限,不对数据包作标记;否则将该包丢弃;

5、如果在此之前，发送过“降低流量”的流量控制信息协议报文，再发送“恢复流量”流量控制信息协议报文。

当该向上游节点发送流量控制信息协议报文的路由器为边缘路由器时，该路由器只向其上游节点发送“降低流量”流量控制信息协议报文。

5 上述平均队列长度的计算依据如下公式：

若队列不空 ($q > 0$)，则： $avg \leftarrow (1 - W_q) \times avg + W_q \times q$

否则： $m \leftarrow f(time - q_time)$

$avg \leftarrow (1 - W_q)^m \times avg$

10 其中： avg = 平均队列长度； $time$ = 当前时间； q_time = 队列空闲时间的起始点； W_q = 队列的权重； q = 队列的实际长度； $f(t)$ = 时间 t 的一个线性函数；

上述概率的计算依据如下公式和步骤：

将 $count$ 加 1；

$P_0 \leftarrow P_{max}(avg - TH_{min}) / (TH_{max} - TH_{min})$

15 $P_s \leftarrow P_0 / (1 - count \times P_0)$

其中： $count$ = 上一个“降低流量”流量控制信息协议报文发送后数据包的分组数

TH_{min} = 队列长度的下限；

TH_{max} = 队列长度的上限；

20 P_0 = 计算中间值；

P_{max} = P_0 的最大值；

P_s = 当前数据包分组进行标注的概率。

路由器流量控制方法包括如下步骤：

25 11、路由器通过“漏桶算法”控制流量；

12、当收到“降低流量”流量控制信息协议报文时，将漏桶容量减少一半；

13、每隔一设定的时间间隔，将漏桶容量加 1 个长度的数据包，但最大值不超过漏桶的最初容量；

14、当收到“恢复流量”流量控制信息协议报文时，立即恢复漏桶的最初容量。

5 其中：步骤 11 所述的“漏桶算法”为如下步骤：

111 以一定的速率产生令牌，令牌桶满时，终止令牌的产生；

112 当数据包到达时，如果令牌桶中有令牌，则取出一个令牌后发送数据包；否则数据包放在缓冲区中等待令牌的出现；

113 当缓冲区满时，丢弃到达的数据包。

10 主机流量控制方法包括如下步骤：

21、主机根据由边缘路由器发来的流量控制信息协议报文中各域的信息查找对应 TCP 连接的套接字 (Socket)；

22、找到对应的 Socket 后，TCP 通过调用支持流量控制信息协议的拥塞控制过程来控制流量；

15 23、如果 TCP 在上一往返时间内没有降低流量，流量控制的请求将在本往返时间内被响应一次；

24、如果该流量控制请求是丢包引起的，TCP 调用原来的拥塞控制过程；

25、如果收到“降低流量”流量控制信息协议报文，TCP 将拥塞窗口和慢启动临界值减半。

20 上述步骤中所述的原来的拥塞控制过程是指以 TCP 为核心的基于窗口技术的端到端的控制；所述的往返时间为一个数据包从一台主机出发到达另一台主机然后返回所用的时间之和；拥塞窗口为 TCP 在启动阶段使用的或在拥塞时期为减小流量而使用的窗口；慢启动临界值为：在慢启动阶段，拥塞窗口的增加方式由以指数增加方式改变为以线性增加方式时的拥塞窗口数量。

25 本发明在数据包队列到达本节点时就开始进行早期的拥塞检测，利用基于随机早期检测的逐节拥塞控制方法和流量控制信息协议报文的传递，采用路由

器流量控制和主机流量控制,使得在网络发生拥塞之前就获得了流量控制;由于拥塞控制是在相邻的节点之间进行,其加快了拥塞控制的响应速度,使丢包率为零,数据延时缩短,延时的抖动减小;且本发明的各节点不需要保存连接的状态信息,具有良好的可扩展性。

5 以下结合附图及一较佳实施例对本发明作进一步的详细说明:

图 1 为本发明一实施例的拥塞控制流程示意图。

图 2 为本发明一实施例的流量控制信息协议报文格式。

图 3 为本发明一实施例以路由器为发送端的拥塞检测控制流程。

图 4 为本发明一实施例以路由器为接收端的接收端拥塞控制流程。

10 图 5 为本发明一实施例以主机为接收端的接收端拥塞控制流程。

如图 1 所示的本发明拥塞控制流程,其包括一发送端主机 1、一接受端主机 4 及其数量不少于两个的路由器 2、3。数据包由发送端 1 发送,经过若干路由器 2、3 的传送后,由接收端 4 接收。在路由器 2 或 3 之中,各有由该路由器 2 或 3 的上一节点发送来的数据包组成的相应的数据包队列 24 或 34。在
15 发送端 1 和接收端 4 中各有其原有的 TCP 拥塞控制程序 11 及 41;而在路由器 2 或 3 之中增设有与之相应的流量控制信息协议报文协议 21 或 31、拥塞检测程序 22 或 32、流量控制程序 23 或 33。

如图 2 所示:流量控制信息协议报文包括如下的信息域:版本号域,其用于验证发送方和接收方的报文格式是否一致;报文长度域,其指出该的报文长度;
20 报文类型域,其指出流量控制信息协议报文的类型为“恢复流量”或“降低流量”;代码域,其指出当前节点的上一节点为主机或路由器;源 IP 地址,其指出发送端主机的 IP 地址;目的 IP 地址,其指出接收端主机的 IP 地址;源端口号,其指出 TCP 连接的发送端主机的端口号;目的端口号,其指出 TCP 连接的接收端主机的端口号;校验域,其用于存放保证报文数据的完整性的校验信息;
25 填充域,其由该报文实际长度与规定的长度值之间的差值构成,用“0”填充。

如图 1 及图 3 所示：当上游节点发送来的数据包到达路由器 2 或 3 时，其拥塞检测程序 22 或 32 将数据包队列 24 或 34 的平均长度与该队列 24 或 34 的平均长度的上限和下限作比较。当该平均长度在上、下限之间时，即认为此时有可能发生拥塞，计算概率 P_0 ，并根据该概率 P_0 对该数据包做标记。对已标记的数据包，路由器 2 或 3 向发送该数据包的上游节点发送“降低流量”流量控制信息协议报文；当平均长度小于下限时，不对该数据包作标记，如果在此之前，已发送过“降低流量”的流量控制信息协议报文，再发送“恢复流量”流量控制信息协议报文；将该数据包加入路由器 2 或 3 的数据包队列 24 或 34 之中；当平均长度大于上限时，丢弃该数据包。

当路由器 2 向其上游节点发送流量控制信息协议报文时，如果该路由器 2 为边缘路由器，则路由器 2 只发送“降低流量”流量控制信息协议报文。

上述平均队列长度的计算依据如下公式：

若队列不空 ($q > 0$)，则：
$$avg \leftarrow (1 - W_q) \times avg + W_q \times q$$

否则：
$$m \leftarrow f(time - q_time)$$

$$avg \leftarrow (1 - W_q)^m \times avg$$

其中： avg = 平均队列长度； $time$ = 当前时间； q_time = 队列空闲时间的起始点； W_q = 队列的权重； q = 队列的实际长度； $f(t)$ = 时间 t 的一个线性函数；

上述的标注概率的计算依据如下公式和步骤：

将 $count$ 加 1；

$$P_0 \leftarrow P_{max} (avg - TH_{min}) / (TH_{max} - TH_{min})$$

$$P_0 \leftarrow P_0 / (1 - count \times P_0)$$

其中： $count$ = 上一个“降低流量”流量控制信息协议报文发送后数据包的分组数

TH_{min} = 队列长度的下限；

TH_{max} = 队列长度的上限；

P_0 = 计算中间值;

P_{max} = P_0 的最大值;

P_s = 当前数据包分组进行标注的概率。

如图 1 及图 4 所示: 路由器 2 或 3 通过“漏桶算法”控制本节点的流量; 当路由器 2 或 3 收到“降低流量”流量控制信息协议报文时, 其将漏桶容量减少一半, 然后每隔一定的时间间隔, 将漏桶容量加 1 个长度的数据包, 但最大值不超过该漏桶的最初容量; 当路由器 2 或 3 收到“恢复流量”流量控制信息协议报文时, 立即恢复该漏桶的最初容量。

所述的“漏桶算法”为: 本节点路由器以一定的速率产生令牌, 当令牌桶满时路由器终止产生令牌; 当一数据包到达本节点路由器时, 如果令牌桶中有令牌, 则取出一个令牌后发送该数据包; 否则将该数据包放在路由器的缓冲区中等待令牌的出现; 当本节点路由器的缓冲区满时, 则丢弃到达本节点路由器的数据包。

- 15 如图 1 及图 5 所示: 发送端 1 的主机根据由边缘路由器 2 发送来的流量控制信息协议报文中各域的信息查找对应 TCP 连接的套接字 (Socket); 如果找到对应的 Socket, 并且, 如果 TCP 在上一往返时间内没有降低流量, 流量控制的请求将在本往返时间内被响应一次; 如果该流量控制请求是由丢包引起的, TCP 则调用原来的拥塞控制过程; 否则, TCP 将拥塞窗口和慢启动临界值
- 20 减半。如果没有找到对应的 Socket, 则报错, 结束。

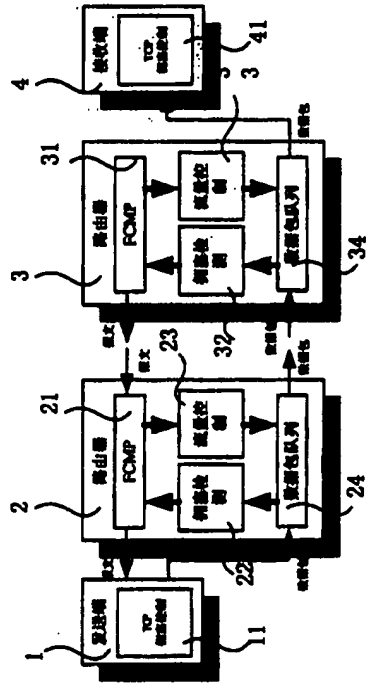


图 1

版本号	报文长度	报文类型	路由器类型代码
源IP地址			
目的IP地址			
源端口号		目的端口号	
校验和		填充域	

图 2

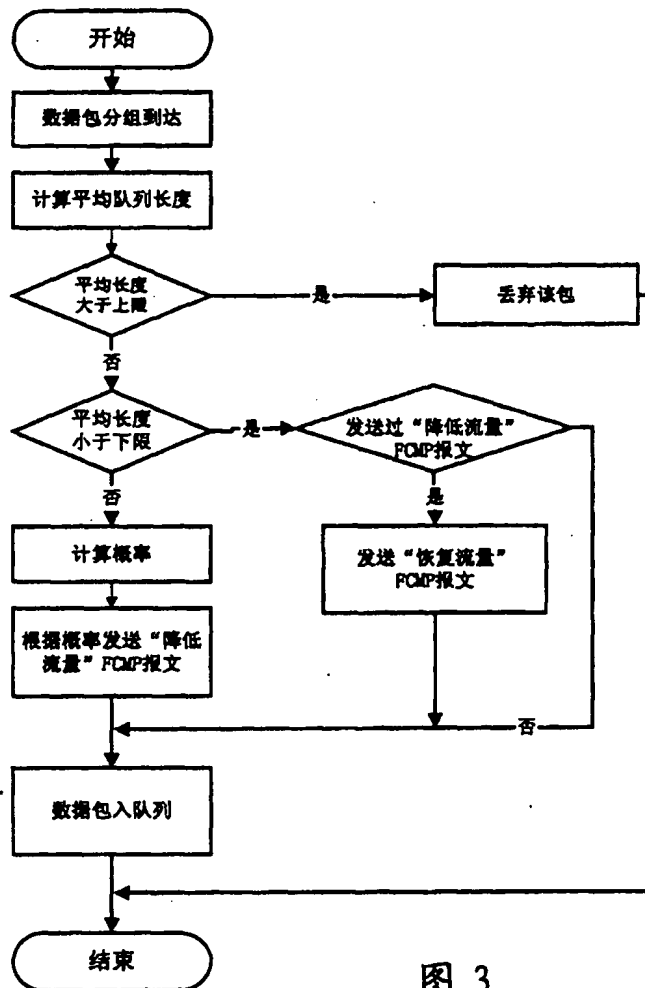


图 3

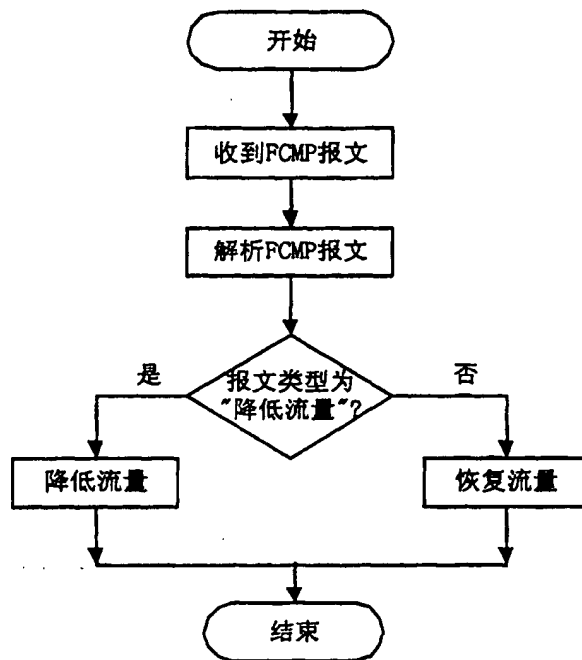


图 4

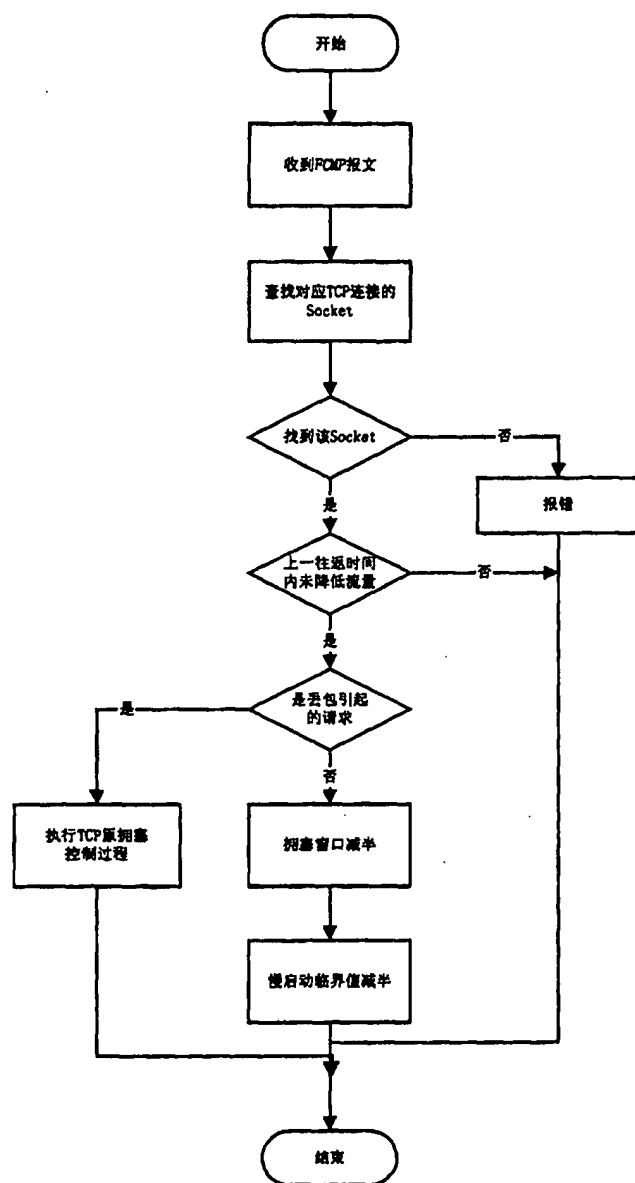


图 5

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.